

Integration of data from multiple studies with
divergent structures and protocols to a coherent
CDISC like data base

David A. Schwab, Ph.D.

SchwabTech

Sept 28, 2010

Introduction

- **Brief history**
- **Database requirements**
- **CDISC conversion**
- **Validation tools**
- **Documentation tool**
- **Integration**
- **Acknowledgements**

Brief History

- **Completed 10 clinical study data integrations to date (11th nearing completion)**
 - **Integrations have ranged from 15 to 29 studies**
- **Supported projects at CROs, small, medium and large Pharmaceutical Companies**
- **Embraced CDISC standards in 2004 and have been using a CDISC-like standard**

Database Requirements

- **Integrity (maintain relationship to source data; not another round of cleaning)**
- **Accurate interpretation (conversion of selection lists, mapping of results)**
- **Simple (updates are not frequent, staff changes can be)**

Conversion to CDISC-like Structure

- **Overall Strategy: keep it simple**
- **Convert to CDISC-like structure (limited data conversion)**
- **Populate relative day variables**
- **Keep 'extra' variables on destination data sets and designate with '_varname'**

Conversion to CDISC-like Structure

- **Programming Strategy**
 - **Bring in all source data**
 - **Combine into 1 data set**
 - **Convert to CDISC structure all at once**
 - **Facilitates debugging because only dealing with one issue at a time: combining source data or translation to CDISC variables**

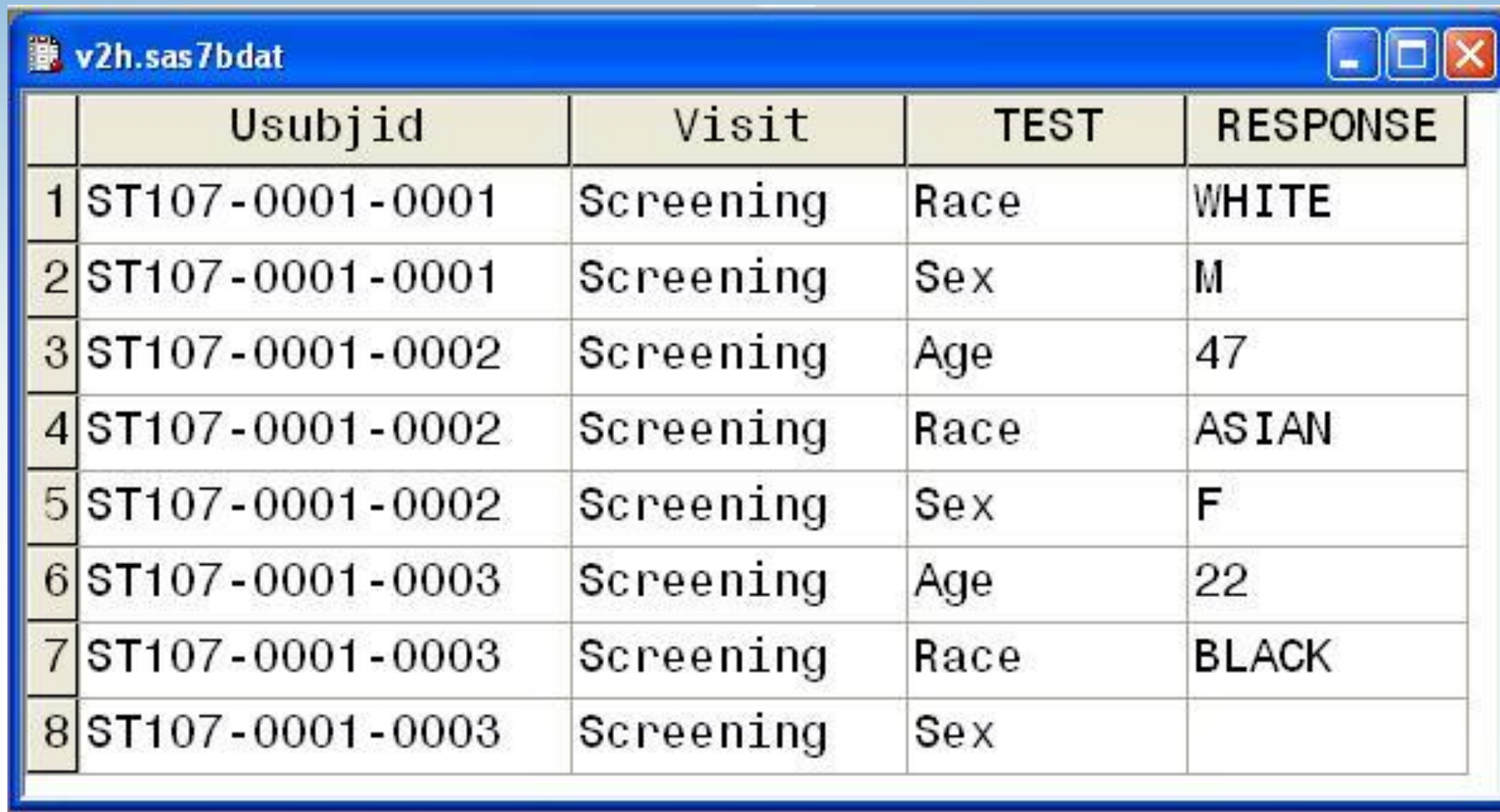
Horizontal to Vertical Structure

- **Bring in source data**
- **Mark data values to be removed**
- **Transform to vertical structure**
- **Remove marked data records**
- **Add CDISC variables**
- **Translate using:**
 - **Direct transfer**
 - **Formats**
 - **Coding (lists or logic)**
- **Validate**
- **Order variables / sort data**
- **Keep CDISC like variables**

Vertical to Horizontal Structure

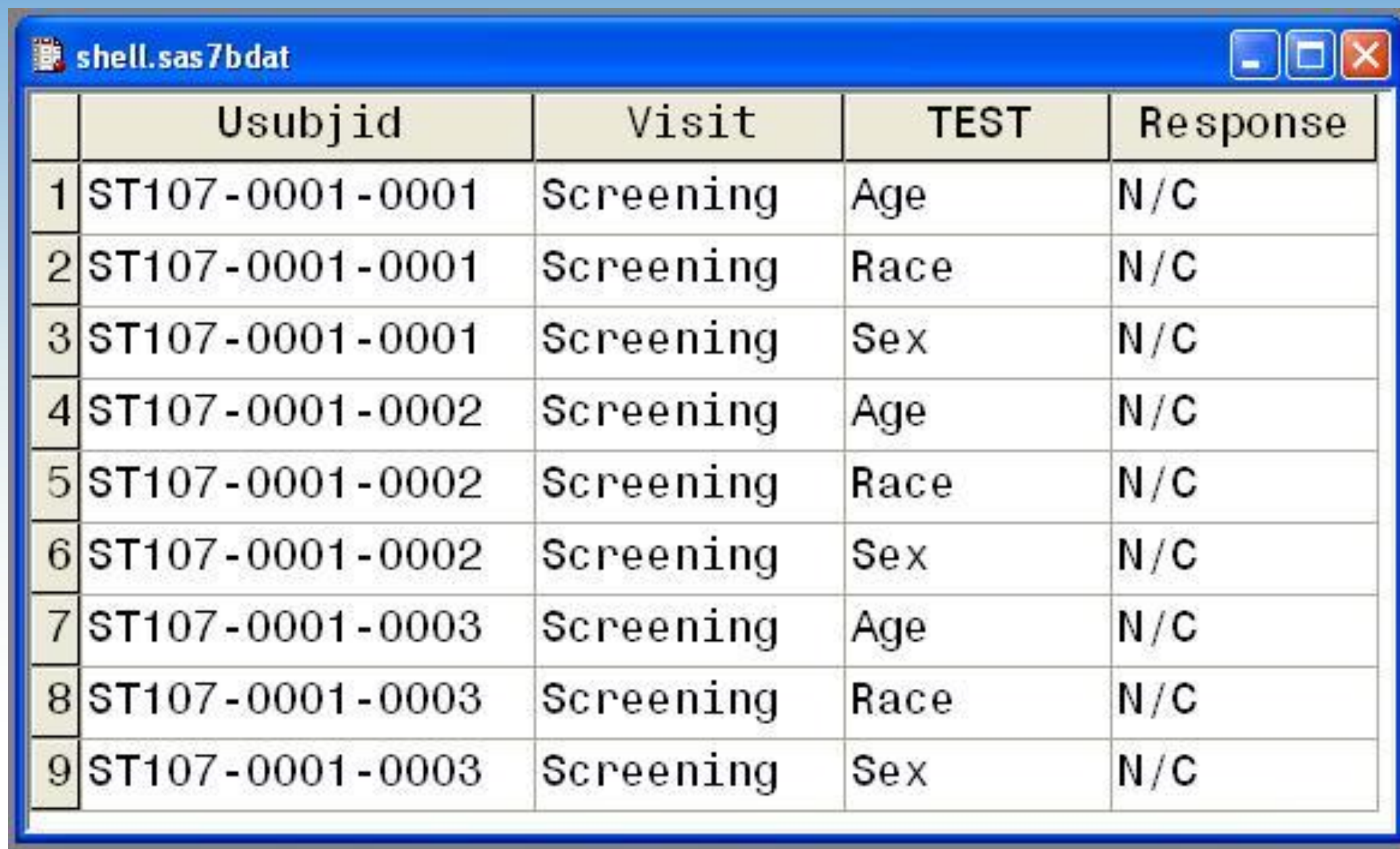
- **Bring in source data**
- **Create data shell: record identifiers x data points**
- **Map source data onto shell (flag data not collected)**
- **Transform to horizontal structure**
- **Add CDISC variables**
- **Translate using:**
 - **Direct transfer**
 - **Formats**
 - **Coding (lists or logic)**
- **Validate**
- **Order variables / sort data**
- **Keep CDISC like variables**

Vertical to Horizontal Structure



	Usubjid	Visit	TEST	RESPONSE
1	ST107-0001-0001	Screening	Race	WHITE
2	ST107-0001-0001	Screening	Sex	M
3	ST107-0001-0002	Screening	Age	47
4	ST107-0001-0002	Screening	Race	ASIAN
5	ST107-0001-0002	Screening	Sex	F
6	ST107-0001-0003	Screening	Age	22
7	ST107-0001-0003	Screening	Race	BLACK
8	ST107-0001-0003	Screening	Sex	

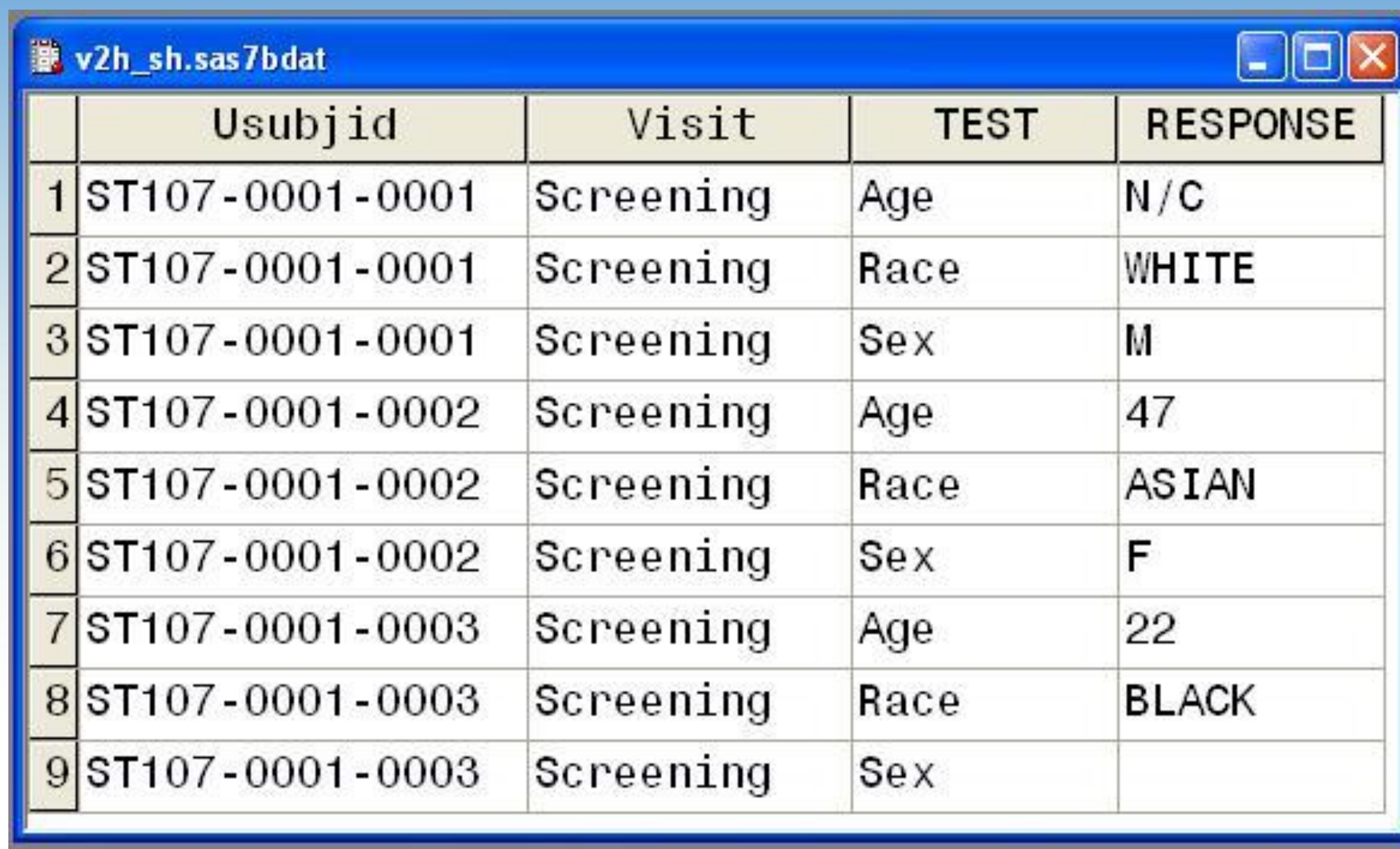
Vertical to Horizontal Structure



The screenshot shows a SAS data window titled "shell.sas7bdat". The window contains a table with 9 rows and 5 columns. The columns are labeled "Row", "Usubjid", "Visit", "TEST", and "Response". The data is organized in a vertical structure, where each row represents a single observation for a specific subject and visit.

	Usubjid	Visit	TEST	Response
1	ST107-0001-0001	Screening	Age	N/C
2	ST107-0001-0001	Screening	Race	N/C
3	ST107-0001-0001	Screening	Sex	N/C
4	ST107-0001-0002	Screening	Age	N/C
5	ST107-0001-0002	Screening	Race	N/C
6	ST107-0001-0002	Screening	Sex	N/C
7	ST107-0001-0003	Screening	Age	N/C
8	ST107-0001-0003	Screening	Race	N/C
9	ST107-0001-0003	Screening	Sex	N/C

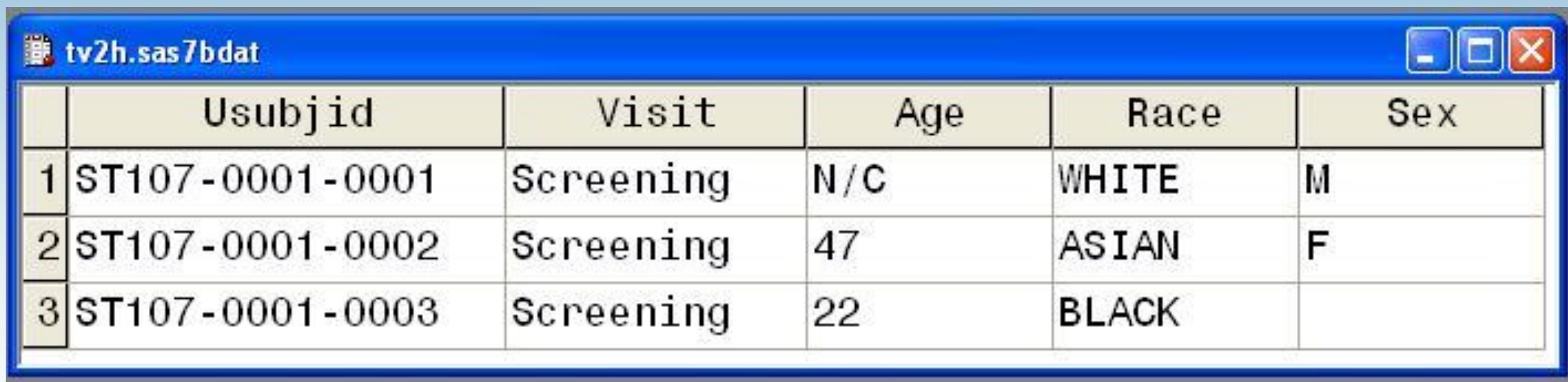
Vertical to Horizontal Structure



The screenshot shows a SAS data window titled "v2h_sh.sas7bdat". The window contains a table with 5 columns: "Row", "Usubjid", "Visit", "TEST", and "RESPONSE". The data is organized in a vertical structure, with each row representing a single observation. The "Usubjid" column contains subject identifiers, the "Visit" column contains the visit type, the "TEST" column contains the test name, and the "RESPONSE" column contains the test result. The "Row" column is a simple index from 1 to 9.

	Usubjid	Visit	TEST	RESPONSE
1	ST107-0001-0001	Screening	Age	N/C
2	ST107-0001-0001	Screening	Race	WHITE
3	ST107-0001-0001	Screening	Sex	M
4	ST107-0001-0002	Screening	Age	47
5	ST107-0001-0002	Screening	Race	ASIAN
6	ST107-0001-0002	Screening	Sex	F
7	ST107-0001-0003	Screening	Age	22
8	ST107-0001-0003	Screening	Race	BLACK
9	ST107-0001-0003	Screening	Sex	

Vertical to Horizontal Structure



	Usubjid	Visit	Age	Race	Sex
1	ST107-0001-0001	Screening	N/C	WHITE	M
2	ST107-0001-0002	Screening	47	ASIAN	F
3	ST107-0001-0003	Screening	22	BLACK	

Validation Tools

- **Simple, quick and easy to interpret = Frequency counts**
 - **Record identifiers (e.g. usubjid, visit, xxTPT)**
 - **Number of Subjects**
 - **Number of Records**
 - **H->V # Records x No Fields = Total # of Records**
 - **V->H Remove duplicates based on Subject, Visit, etc.**
 - **Data Fields: Frequency by Values**
- **Direct Comparison**
 - **Comprehensive comparison using 'long way home' macros**

Validation Tools (ex)

source.irritation - Subject Count

<u>STUDYNO</u>	<u>Frequency</u>
Boston-928	774

edisc.sa - Subject Count

<u>STUDYID</u>	<u>Frequency</u>
PHUSE/BOSTON-928	774

Validation Tools (ex)

source.irritation - Record by Visit Count

<u>VISIT</u>	<u>Frequency</u>
--------------	------------------

2	769
---	-----

3	629
---	-----

4	569
---	-----

5	744
---	-----

edisc.sa - Record by Visit Count

<u>VISITNUM</u>	<u>VISIT</u>	<u>Frequency</u>
-----------------	--------------	------------------

2	VISIT B (BASELINE)	769
---	--------------------	-----

3	VISIT 4C	629
---	----------	-----

4	VISIT 8C	569
---	----------	-----

5	VISIT 12C (FINAL)	744
---	-------------------	-----

Validation Tools (ex)

source.irritation – treated arm

ARMTRT Frequency

RIGHT ARM	1546
LEFT ARM	1163

edisc.sa - Original Result

(where satestcd='ARMTRT')

Original Result

SAORRES Frequency

RIGHT ARM	1546
LEFT ARM	1163

Validation Tools

Direct Comparison Macros

- **H2V Macro**
 - Method: uses multiple output statements
- **V2H Macro**
 - Method: retain multiple variables
 - Use first.var to reset values
 - and last.var to collapse by outputting

Validation Tools

- **Frequency Counts**
 - **Record identifiers**
 - **Usubjid, visit, xxTPT**
 - **Transformed or mapped fields**
- **Overall direct comparison**

Documentation Tool

- **Tagged comments within programs**
- **Easy to document within programs**
- **Give programmers feedback**

Documentation Tool (pgm)

```
*#irritation/data sets/sa;  
data sa;  
set sa;
```

```
*#/set to 'ST-03-111' per protocol/studyid*;  
STUDYID="&study";
```

```
SITEID='00' || substr(patno,1,2);
```

```
*# patno/subset on last 3 digits to get subjid / subjid *;  
SUBJID='0' || substr(patno,3);
```

```
*# /Studyid+siteid+subjid/usubjid *;  
USUBJID=trim(left(STUDYID)) || '-' || trim(left(siteid)) || '-'  
|| trim(left(subjid));
```

Documentation Tool

- **Implementation**
 - **Read source data set structures**
 - **Read destination CDISC data set structures**
 - **Parse out comments from programs**
 - **Link source to destination variables using comments**
 - **Account for all variables in source and destination data sets**

Documentation Tool (cnv)

Raw Data Set(s): IRRITATION			eDISC Data Set: SA	
Raw Variables			eDISC Variables	
Label	Name	Mapping Comment	Name	Label
		Set to 'ST-03-111' per protocol	STUDYID	Study Identifier
		Set to SA (CDISC standard)	DOMAIN	Domain Abbreviation
			SITEID	Study Site Identifier
PATIENT NUMBER	PATNO	Subset on last 3 digits to get subjid	SUBJID	Subject Identifier for the Study
		Studyid+siteid+subjid	USUBJID	Unique Subject Identifier
		Dynamically set based on subj records	SASEQ	Sequence Number
VISIT	VISIT	Decode of visit using vis. format	VISIT	Visit Name
VISIT	VISIT	Decode of visit using visnum. format	VISITNUM	Visit Number
VISIT	VISIT	Decode of visit using visitdy. format	VISITDY	Planned Study Day of Visit
VISIT DATE	VISIT_DT	Reformat visit date to yymmdd10.	SADTC	Date/Time of Collection
		Spid defined by satestcd lookup table	SASPID	Sponsor-Defined Identifier
TREATED ARM	ARMTRT	From variable name	SATESTCD	Question Short Name
TREATED ARM	ARMTRT	From variable label	SATEST	Question
		Set to STUDY ARM	SACAT	Category
			SASCAT	Subcategory
TREATED ARM	ARMTRT	Decode of armtrt using arm. format	SAORRES	Original Result
		Direct transfer from saorres	SASTRESC	Result in Std Format
		Status and reason nd not collected	SASTAT	Completion Status
		Status and reason nd not collected	SAREASND	Reason Not Done
		No comment variable present	SACOMM	Comments
		Sady defined as sadtc-dm.rfstdtc+1	SADY	Study Day of Collection
STUDY NUMBER	STUDYNO	Not needed, not mapped		

Integration

- **Simple as stacking**
- **Coding of Adverse Events and Concomitant Medications**
- **Standardize lists (integrate literals into selections)**
- **Implement standardization of meta-data (xxTEST and xxTESTCD)**
- **Establish final xxSEQ variable values**

Integration (2)

- Tag screening, baseline and final visits
- Develop relative day – relative day is key
 - Relative or integrated visit
 - Develop visit numbers
- Move _varname to Suppqual using usubjid, dataset name and xxSEQ value
- Move _testcd to suppqual using usubjid and xxgrpid or xxdtc

Trials Combined with Integrated Visit (2)

Type	Integrated Visit / Number		Relative Day	Trial Visit / Number	
Trial 1	Week -2	-2	-8	Screening	-99
	Baseline	1	1	Baseline	1
	Week 5	5	32	Visit 1	2
	Week 9	9	60	Final (Visit 2)	3
Trial 2	Week -4	-4	-19	Screening	-99
	Week -2	-2	-9	Washout	-1
	Baseline	1	1	Baseline	1
	Week 2	2	18	Visit 1	2
	Week 5	5	32	Final (Visit 2)	3
Trial 3	Week -7	-7	-39	Screening	-99
	Week -5	-5	-28	Washout	-1
	Week -4	-4	-19	Titration	0
	Week -2	-2	-9	Titration	0
	Baseline	1	1	Baseline	1
	Week 3	3	23	Visit 1	2
	Week 7	7	44	Final (Visit 2)	3

Trials Combined with Integrated Visit

Integrated Visit / Number		Trial 1 Visit / Number		Trial 2 Visit / Number		Trial 3 Visit / Number	
Week -7	-7					Screening	-99
Week -5	-5					Washout	-1
Week -4	-4			Screening	-99	Titration	0
Week -2	-2	Screening	-99	Washout	-1	Titration	0
Baseline	1	Baseline	1	Baseline	1	Baseline	1
Week 2	2			Visit 1	2		
Week 3	3					Visit 1	2
Week 5	5	Visit 1	2	Final (Visit 2)	3		
Week 7	7					Final (Visit 2)	3
Week 9	9	Final (Visit 2)	3				

Trials Combined: Safety Analysis

Where trial visit number in (-1,0,1)

Type	Integrated Visit / Number		Relative Day	Trial Visit / Number	
Trial 1	Week -2	-2	-8	Screening	-99
	Baseline	1	1	Baseline	1
	Week 5	5	32	Visit 1	2
	Week 9	9	60	Final (Visit 2)	3
Trial 2	Week -4	-4	-19	Screening	-99
	Week -2	-2	-9	Washout	-1
	Baseline	1	1	Baseline	1
	Week 2	2	18	Visit 1	2
	Week 5	5	32	Final (Visit 2)	3
Trial 3	Week -7	-7	-39	Screening	-99
	Week -5	-5	-28	Washout	-1
	Week -4	-4	-19	Titration	0
	Week -2	-2	-9	Titration	0
	Baseline	1	1	Baseline	1
	Week 3	3	23	Visit 1	2
	Week 7	7	44	Final (Visit 2)	3

Trials Combined: Efficacy Analysis

Where trial visit number in (0,1)

Type	Integrated Visit / Number		Relative Day	Trial Visit / Number	
Trial 1	Week -2	-2	-8	Screening	-99
	Baseline	1	1	Baseline	1
	Week 5	5	32	Visit 1	2
	Week 9	9	60	Final (Visit 2)	3
Trial 2	Week -4	-4	-19	Screening	-99
	Week -2	-2	-9	Washout	-1
	Baseline	1	1	Baseline	1
	Week 2	2	18	Visit 1	2
	Week 5	5	32	Final (Visit 2)	3
Trial 3	Week -7	-7	-39	Screening	-99
	Week -5	-5	-28	Washout	-1
	Week -4	-4	-19	Titration	0
	Week -2	-2	-9	Titration	0
	Baseline	1	1	Baseline	1
	Week 3	3	23	Visit 1	2
	Week 7	7	44	Final (Visit 2)	3

Integration Summary

- Tag screening, baseline and final visits
- Develop relative day – relative day is key
- Use consistent visit numbers to aid in selection of relative day
- Simple = flexible

Data Base Longevity

- **Maintain a solid relationship with source data**
- **Any discrepancies should be noted in the data**
- **Keep integration simple to accommodate future updates**

Acknowledgements

- **Deborah N. Beville**
 - **Director of Data Management**
- **Jagadish Kumar**
 - **Clinical SAS Programmer**
- **Maria Overdorf**
 - **Director of Data Management**

Acknowledgements (2)



The Schwabs (Ashley, Debbie and Amanda)

David A. Schwab, Ph.D.

SchwabTech

“Your success is our success”

**1021 Welsh Ayres Way
Downingtown, PA 19335
Ph: 484.574.3154
Fax: 484.348.2233**

Email: David.Schwab@SchwabTech.com

Website: www.SchwabTech.com